

Desarrollo de Herramientas Inteligentes para la Web Semántica

Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Laboratorio de Investigación y Desarrollo en Ingeniería de Software
Laboratorio de Investigación en Visualización y Computación Gráfica
Laboratorio de Investigación y Desarrollo en Sistemas Distribuidos

Departamento de Ciencias e Ingeniería de la Computación
Instituto de Investigación en Ciencia y Tecnología Informática (IICyTI)
Universidad Nacional del Sur
dcc@cs.uns.edu.ar

El objetivo general de este proyecto, en el marco de las recomendaciones del *World Wide Web Consortium* (W3C), es el desarrollo de las herramientas necesarias para la creación de la infraestructura de soporte cognitivo y la explotación por parte de agentes autónomos del conocimiento almacenado en la Web. Esto incluye lenguajes de representación de conocimiento, herramientas para la creación, mantenimiento y visualización interactiva de ontologías, y máquinas de inferencia especializadas que puedan razonar con ellas y también colaborar en su visualización inteligente. El grupo responsable de este proyecto, creado en el Departamento de Ciencias e Ingeniería de la Computación de la UNS, está integrado por: Guillermo Simari, Alejandro García, Fernando Tohmé, Telma Delladio, Diego Martínez y Sergio Gómez (Laboratorio de Investigación y Desarrollo en Inteligencia Artificial); Elsa Estévez y Pablo Fillottrani (Laboratorio de Investigación y Desarrollo en Ingeniería de Software); Silvia Castro y Sergio Martig (Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica) y por Jorge Ardenghi y Javier Echaiz (Laboratorio Investigación y Desarrollo en Sistemas Distribuidos).

Introducción

La World Wide Web ha cambiado en forma dramática la disponibilidad de información en formato electrónico. La cantidad de documentos en línea crece aceleradamente habiendo llegado a algunos miles de millones durante el corriente año. Al mismo tiempo, varios cientos de millones de usuarios realizan búsquedas sobre este repositorio, y acceden y contribuyen al mismo en forma continua. No es sorprendente entonces que estas tareas sean cada vez más dificultosas y menos efectivas.

La mayor parte del contenido de la Web está diseñado, naturalmente, para que sea inteligible para los seres humanos. Por otro lado, en la actualidad los programas que deben manipularla carecen de las capacidades para interpretar el significado (la semántica) de su contenido. El reconocimiento de este problema llevó a Tim Berners-Lee, creador de la World Wide Web, a proponer una nueva forma de contenido en la Web. Este contenido deberá ser significativo para los propios sistemas computacionales, *i.e.* deberá ser interpretable por ellos. La nueva estructura de la Web, creará un entorno en el que los agentes de software puedan recorrer las páginas *entendiendo* su significado. Esta extensión de la Web, denominada por Berners-Lee “*Web Semántica*”, provee a la Web de significado *bien definido*, lo que facilitará la cooperación entre seres humanos y agentes de software, dándole a estos últimos la habilidad de *comprender* el contenido de la Web, razonar con él y así aumentar su efectividad.

En octubre de 1994 se creó el *World Wide Web Consortium* [W3C] con el objetivo de llevar a la World Wide Web hacia el total desarrollo de su potencial introduciendo protocolos comunes que promuevan su evolución y garanticen su interoperabilidad. Para poder responder a las expectativas crecientes de los usuarios de la Web, el consorcio W3C está diseñando la próxima generación de la Web proponiendo las tecnologías que ayudarán a hacer que ésta adquiera una estructura robusta, escalable y adaptiva para el manejo de la información.

Las metas a largo plazo del W3C son: proveer el *acceso universal*, el desarrollo de la *Web Semántica* y el desarrollo de la llamada *Web of Trust*, i.e. una guía para la consideración de las cuestiones legales, comerciales y sociales que este desarrollo tecnológico ha introducido. Los principios fundamentales de diseño son: *Interoperabilidad* provista a través de especificaciones de lenguajes y protocolos de la Web, *Simplicidad* y *Modularidad* para dar cabida a la *Evolución* permitiendo que las nuevas tecnologías puedan introducirse con mayor facilidad, *Arquitectura Descentralizada* de manera que el crecimiento de la Web pueda realizarse sin la dependencia de estructuras centralizadas.

Relevancia del Problema

La Web Semántica es un esfuerzo para dotar a la Web con conocimiento embebido que pueda ser accedido por las aplicaciones. El resultado esperado es introducir la posibilidad de que las páginas no sólo puedan ser consultadas por cadenas de caracteres como en la actualidad, sino que puedan ser consultadas por el conocimiento que almacenan. Esta extensión de la Web dará la oportunidad para el desarrollo de aplicaciones, tales como los agentes autónomos, que exploten de manera intensiva la posibilidad de acceder a ese conocimiento.

La Web Semántica se encuentra en las primeras etapas de construcción. Es por esta razón que *es de extrema importancia en este momento desarrollar tecnología propietaria cuando la oportunidad está abierta*.

La Web no sólo es inmensa, sino que también es dinámica, evolutiva, potencialmente inconsistente, y en general las respuestas obtenidas son dependientes del contexto. Como nuevo espacio de desarrollo, existe un potencial enorme para realizar en este momento aportes innovadores que permitan el uso eficiente de la Web para aplicaciones aprovechables en todos los sectores: institucionales educativos, de salud, gubernamentales y empresariales de todo nivel. La posibilidad de desarrollar Servicios de Web que puedan explotar la información semántica apropiada, hará accesible la información por medio de aplicaciones autónomas y contribuirá al desarrollo de contenidos que permitirán la difusión de información de todo tipo en forma mucho más efectiva. El uso de estándares y recomendaciones ofrecidos por la W3C permitirá realizar la investigación y el desarrollo de las herramientas propuestas como software de código abierto y compatible con otros desarrollos. La construcción de ontologías, con la contribución desde el sector del conocimiento, aceptadas por el sector público y privado llevará a una mejor vinculación entre todos los actores permitiendo una articulación fluida de los desarrollos particulares.

Estas características hacen de la Web Semántica un área de investigación de gran importancia. Casi todas las áreas teórico-prácticas en Ciencias e Ingeniería de la Computación deberán aportar contribuciones significativas para lograr avances importantes. En particular, la Inteligencia Artificial, la Computación Gráfica y la Visualización, la Ingeniería del Software y los Sistemas Distribuidos, son de particular interés para este proyecto.

Objetivos Específicos

La tarea de los investigadores es la de agregar capacidades de representación de conocimiento y razonamiento a la red. Esto significa proveer a los agentes de software con la habilidad de reconocer la representación, de usar reglas para realizar inferencias y de planear acciones en función de metas propias y requeridas por sus usuarios. Es posible desarrollar los elementos necesarios para estas tareas en el marco de XML (eXtended Markup Language) y RDF (Resource Description Framework). El lenguaje XML da la posibilidad de agregar estructura a los documentos sin darles significado. Este significado es introducido por los elementos de RDF que permiten expresar relaciones entre elementos en la red que resultan individualizados por medio de URIs (Uniform Resource Locators).

Este marco se completa con la creación de *Ontologías*. Una ontología, desde el punto de vista de la Inteligencia Artificial, define las *palabras y conceptos* (o *significados*) utilizados para describir y representar un área del conocimiento, estandarizando de esa manera los significados. El uso de ontologías permite definir con claridad *qué existe* de manera única, facilitando el uso de la información en el razonamiento y la acción por medio de los agentes de software.

En este contexto, el objetivo de la visualización es ayudar a encontrar rápidamente la información semántica relevante y a crear, modificar y explorar su estructura fácilmente. La *Visualización* permite mejorar la percepción de la estructura de grandes espacios de información por parte del usuario y provee distintas facilidades de interacción. Es necesario contar con técnicas de visualización que se constituyan en herramientas para visualizar ontologías. Éstas deben brindar la posibilidad de visualizar no sólo los esquemas de las ontologías y las instancias relacionadas sino también su integración de una manera efectiva. En este sentido se hace clara la necesidad de contar con herramientas específicamente diseñadas que permitan visualizar este tipo de ontologías con una gran cantidad de instancias explotando la naturaleza jerárquica y solapada de su jerarquía de clases y posibilitando así la realización de una variedad de tareas según el tipo de usuario y el estadio del ciclo de vida de la ontología en el que se esté trabajando.

El *Desarrollo Basado en Componentes (DBC)* es una metodología para construir sistemas de software complejos, que enfatiza la construcción de sistemas mediante el ensamblaje de bloques funcionales existentes. Por otro lado, el desarrollo de software orientado a agentes extiende el desarrollo tradicional de software basado en componentes. Los componentes encapsulan agentes inteligentes, facilitando la comunicación entre ellos, la integración y la flexibilidad. De este modo, la tecnología basada en esta clase de componentes resulta la más adecuada para el desarrollo de herramientas de software inteligentes que descubran, utilicen, y visualicen información suministrada. Las ventajas provistas por el uso de componentes de software, por el desarrollo basado en componentes, y por arquitecturas específicas para un dominio de aplicación son especialmente adecuadas en aplicaciones científicas de alta performance, complejas y que generalmente requieren la combinación de la experiencia de distintos grupos de investigación.

En particular, debido a la complejidad del software a construir se deberán asociar componentes de muy diversos niveles de complejidad y funcionalidad. Por ejemplo, habrá componentes específicamente dedicados a definir la semántica del modelo, otros que proveerán servicios inteligentes para la realización de las búsquedas, y selección de información, y otros componentes responsables de proveer los servicios de visualización. Por este motivo, es de vital importancia la definición de los frameworks y las arquitecturas adecuados para cada uno de los dominios específicos [AHLM03, BCK98, CKK02, DKW01]. Para cumplir con este objetivo, se estudiarán y analizarán distintas arquitecturas y frameworks en base a los servicios que deberán proveer los distintos tipos de componentes. En este caso, el aporte del trabajo de investigación consiste en definir nuevos frameworks y arquitecturas para los diferentes dominios que abarca la explotación de la Web Semántica.

La arquitectura a desarrollar deberá enfatizar que sus subestructuras se puedan entender y modificar fácilmente. Asimismo, deberá integrar los aspectos específicos de inteligencia artificial en lo referente a los componentes inteligentes, responsables de seleccionar los distintos sitios de interés para la búsqueda requerida, con facilidades de recuperación de documentos, y con las herramientas de búsqueda, selección, y visualización de la información solicitada. De igual modo, deberá satisfacer todos los atributos de calidad necesarios para los sistema de búsqueda, recuperación, y visualización de información, como por ejemplo performance y facilidad de uso entre otros.

En el desarrollo del proyecto se buscarán resultados teórico-prácticos que extiendan el trabajo previo de los investigadores integrantes del Grupo. En el área de Representación de Conocimiento y Razonamiento se dispone del desarrollo de la máquina de inferencia para el razonamiento rebatible, Defeasible Logic Programming (DeLP) [GS04], y se extenderán estos desarrollos para incluir la noción de utilidad en la máquina de inferencia básica de DeLP. El desarrollo de los formalismos de razonamiento rebatible estuvo en un principio fundamentado en analogías con procesos de decisión, estas similitudes dieron lugar a la aplicación del razonamiento rebatible a la formalización de procesos de negociación. El razonamiento rebatible puede dar una forma de determinar las utilidades de un conjunto de opciones. Estas utilidades no son numéricas, sino simplemente en forma de orden de preferencias. Se explorará la introducción de una caracterización general de "utilidad de un argumento" y su uso en el marco de la DeLP. Una consecuencia inmediata de esta variante del razonamiento rebatible va estar en su aplicación a problemas de planeamiento, en los que un agente inteligente tiene que comparar entre distintos planes potenciales de acuerdo a criterios tales como el mejor uso de los recursos disponibles, menor riesgo, *etc.* Finalmente, esta introducción de utilidades en la DeLP permitirá también resolver el problema de la formación de creencias por parte de agentes racionales, que tan importante es en el área del razonamiento bayesiano. De hecho, la introducción de utilidades en la actualización rebatible de creencias [FKS02] permitirá un tratamiento no-numérico del razonamiento con incertidumbre y de la dinámica racional del conocimiento.

Desde la perspectiva de la Visualización, se buscarán metáforas adecuadas de representación e interacción y se explorarán tales metáforas para poder analizar su aplicabilidad en este contexto. Se analizarán las representaciones visuales existentes en el campo de Visualización de Información para su aplicabilidad en el contexto de Visualización de Ontologías en la Web Semántica. Se explorarán representaciones visuales alternativas y se determinarán las interacciones requeridas por cada técnica para lograr una interacción efectiva con la Web Semántica.

El aporte desde la Ingeniería de Software será definir arquitecturas de software basadas en componentes adecuados para cada una de las herramientas. Se definirán frameworks de componentes específicos para los distintos dominios de aplicación y métricas para evaluar atributos internos y externos de frameworks y arquitecturas. Se propondrá la arquitectura y los frameworks adecuados para cada una de las herramientas.

Bibliografía

[AHLM03] Apperly, H., Hofman, R., Latchem, S., Maybank, B., McGibbon, B., Piper, D., Simons, Ch., *Service -and Component- Based Development, Using the Select Perspective and UML*, Addison Wesley, 2003. [BB95] Baecker, R. M., Buxton, W. A., *Readings in Human-Computer Interaction*. San Mateo CA.: Morgan Kaufmann Publishers, 1995.

[BYRN99] Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison Wesley, 1999.

[BCK98] Bass, L., Clements, P., Kazman, R., *Software Architecture in Practice*, Addison Wesley, 1998.

- [BS03] Bederson, B., Shneiderman, B., The Craft of Information Visualization: Readings and Reflections, Morgan Kaufmann Pub, 2003.
- [BL02] Berners-Lee, T. Weaving the Web. Harper, San Francisco, 2002.
- [BLHL01] Berners-Lee, T., Hendler, J., Lassila, O., The Semantic Web. Scientific American, volume 284, number 5, mayo 2001, págs.34-43.
- [BL85] Brachman, R.J., Levesque, H. Readings in Knowledge Representation. Morgan Kaufmann, 1985.
- [CKK02] Clements, P., Kazman, R., Klein, M., Evaluating Software Architectures, Methods and Case Studies, Addison Wesley, 2002.
- [CMS99] Card, S., Mackinlay, J., Shneiderman, B., Readings in Information Visualization – Using Vision to Think, Morgan Kaufmann, 1999.
- [DAML01a] DAML-S, <http://www.daml.org/services>, 2001.
- [DAML01b] DAML-S Coalition: A.Ankolekar, M.Burstein, J.Hobbs, O.Lassila, D.Martin, S.McIlraith, S.Narayanan, M.Paolucci, T.Payne, K.Sycara, y H.Zeng. DAML-S: Semantic markup for Web services. Proc. Int.Semantic Web Working Symp. (SWWS), 411-430, 2001.
- [DCvH02] Dean, M., Connolly, D., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L., OWL, Web Ontology Language 1.0 Reference W3C Working Draft, 29 Julio de 2002. Disponible en <http://www.w3.org/TR/owl-ref>.
- [DKW01] Dikel, D.; Kane, D., Wilson, J., Software Architecture. Organizational Principles and Patterns, Prentice Hall Inc., 2001.
- [DFAB98] Dix, A., Finlay, J., Abowd, G., Beale, R., Human-Computer Interaction, Prentice Hall Europe, Second Edition, 1998.
- [DIB99] G. Di Battista, P. Eades, R. Tamassia y I. Tollis, Graph Drawing, Algorithms for the visualization of graphs, Prentice Hall, 1999.
- [FGW02] Fayyad, U., Grinstein, G., Wierse, A., Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann Pub., 2002.
- [FKS02] Falappa, M.A., Kern-Isberner, G., Simari, G.R. Explanations, Belief Revision and Defeasible Reasoning. Artificial Intelligence Journal, 141: pp. 1-28, 2002.
- [GCh03] Fayyad, U., Grinstein, G., Wierse, A., Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann Pub., 2002.
- [GF02] Geroimenko, V., Chen, Ch., (Eds.) Visualizing the Semantic WEB, Springer Verlag, 2003, Second Print.
- [GS04] García,A.J., Simari,G.R. Defeasible Logic Programming: An Argumentative Approach. Theory and Practice of Logic Programming. Vol 4(1-2) pp 95-138, 2004.
- [KSL93] Towards Principles for the Design of Ontologies Used for Knowledge Sharing, Technical Report KSL - 93-04. Knowledge Systems Laboratory, Stanford University, 1993.
- [Lab96] Labrou, Y. Semantics for an Agent Communication Language. Ph.D. Thesis, University of Maryland, 1996.
- [Shn98] Shneiderman, B., Designing the User Interface, Addison-Wesley Pub. Company, 1998.
- [Spe01] Spence, R., Information Visualization, Addison-Wesley, ACM Press, 2001.
- [Tuf90] Tufte, E.R., Envisioning Information, Cheshire, CT Graphics Press, 1990.
- [Tuf97] Tufte, E.R., Visual Explanations: Images and Quantities, Evidence and Narrative, Cheshire, CT Graphics Press, 1997.
- [W3C] Documentos y recomendaciones del sitio <http://www.w3.org> (World Wide Web Consortium).